DØ Note 6053-CONF

Measurement of the $\sigma(p\bar{p} \to Z + b\,\text{jet})/\sigma(p\bar{p} \to Z + \text{jet})$ Ratio in Dilepton Final States Using 4.2 fb^{-1} of Data Collected with the DØ Detector at $\sqrt{s} = 1.96$ TeV

The DØ Collaboration
URL http://www-d0.fnal.gov
(Dated: April 20, 2010)

The inclusive production of b jets in association with a Z boson provides an important test of QCD predictions. This paper describes a measurement of the ratio of inclusive cross sections $\sigma(p\bar{p} \to Z + b\,\text{jet})/\sigma(\text{p}\bar{\text{p}} \to \text{Z} + \text{jet})$ when the Z boson decays into electrons or muons. The b jets are identified and separated from c and light jets by means of a neural network jet tagging algorithm that exploits a longer mean lifetime of heavy flavor hadrons relative to the light ones. The measurements use a data sample of 4.2 fb^{-1} of $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV collected by the DØ detector. The measured ratio $\sigma(Z + b\,\text{jet})/\sigma(\text{Z} + \text{jet})$ is 0.0176 $\pm$ 0.0024 (stat.) $\pm$ 0.0023 (syst.) for jets with transverse momenta $p_T > 20$ GeV and pseudorapidities $|\eta| \leq 1$. The results are consistent with perturbative QCD calculations at next-to-leading order.

Preliminary Results for the 2010 Conferences

I. INTRODUCTION

The measurement of the Z boson production cross section in association with one or more b jets provides an important test of perturbative quantum chromodynamics (QCD) calculations [1]. The understanding of this process and its description by current theoretical calculations is important because it provides a major background to various analyses, e.g. searches for the Standard Model Higgs boson in the $Z(\to \ell^+\ell^-/\nu\bar{\nu})H(\to b\bar{b})$ associated production and to the searches for supersymmetric partners of b quarks. The process is also sensitive to the b quark density in the proton, which is usually derived perturbatively from the gluon component of the parton distribution functions (PDF). A precise knowledge of the b quark density is necessary to accurately predict processes that strongly depend on it, such as electroweak production of a single top quark and production of non-standard model Higgs bosons in association with b quarks.

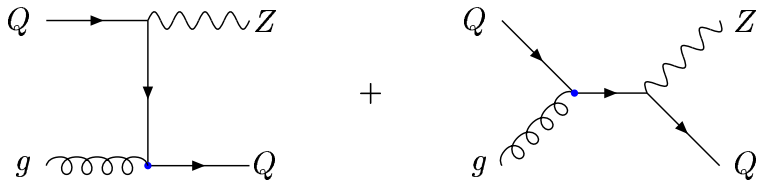


FIG. 1: Diagrams contributing to the $Qg \to ZQ$ ($Q = b, c$) production.

FIG. 2: Diagrams contributing to the $q\bar{q} \to ZQ\bar{Q}$ production.

The parton-level subprocesses that contribute to $Z + b$ jet final states are shown in Figs. 1 and 2. In Fig. 1, a single b quark from the proton undergoes a hard scatter and the other b quark typically remains soft, forward and therefore undetected. In $q\bar{q}$ initiated processes, Fig. 2, a $b\bar{b}$ pair may be reconstructed in the same jet, or one of the b quarks is not accepted by the detector or analysis selection criteria. Theoretical calculations for $Z + b$ production are currently available at the next-to-leading order (NLO) in QCD [1, 2]. According to perturbative QCD calculations, both types of subprocesses are expected to contribute almost equally to the $Z + b$ jet production at the Tevatron.

The DØ experiment provided the first measurement of the ratio of cross sections for $Z + b$ jet to $Z+$ jets production. The measured value of the ratio was found to be 0.021 ± 0.005 [3], which is consistent with the theoretical prediction of 0.018 ± 0.004 [1]. This result was based on 180 pb^{-1} of data and the analysis assumed the ratio of the $Z + b$ jet to $Z + c$ jets cross section from NLO calculations. The CDF experiment has also measured the ratio, as well as the $Z + b$ jet inclusive cross section using 2 fb^{-1} of data [4]. Their analysis did not use the theoretical prediction for $\sigma(Z + c \text{ jet})/\sigma(\text{Z} + \text{b jet})$ and extracted the fractions of different jet flavors from data. The ratio of $Z + b$ jet to inclusive $Z+$ jets production measured by CDF is $0.0208 \pm 0.0033\,(\text{stat.}) \pm 0.0034\,(\text{syst.})$.

This paper describes a measurement of $\sigma(Z + b \text{ jet})/\sigma(\text{Z} + \text{jet})$ when the Z boson decays into dilepton final states (pairs of muons or electrons) in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV. The measurement of the ratio benefits from the cancellations of many systematic uncertainties such as uncertainty from luminosity and those related to the identification of leptons and jets. The remaining systematic uncertainties arise from the difference between b jet and c and light jet properties.

II. DATA SAMPLE AND EVENT SELECTION

The DØ detector [5] has a central-tracking system [6] consisting of a silicon microstrip tracker (SMT) and a central fiber tracker located within a 2 T superconducting solenoidal magnet. The tracking coverage extends to a pseudorapidity of $|\eta_{\rm det}| \lesssim 3$; $\eta_{\rm det}$ is measured with respect to the center of the DØ detector, while η without subscript represents the physical pseudorapidity. A liquid-argon and uranium calorimeter has a central section (CC) covering $|\eta_{\rm det}|$ up to ≈ 1.1, and two end calorimeters (EC) that extend coverage to $|\eta_{\rm det}| \approx 4.2$ [5]. The region $1.1 < |\eta_{\rm det}| < 1.5$ is called the inter-cryostat region (ICR) and has limited coverage by the electromagnetic calorimeter. An outer muon system, at $|\eta_{\rm det}| \lesssim 2$, consists of a layer of tracking detectors and scintillation trigger counters in front of 1.8 T toroids, followed by two similar muon detector layers outside of the toroids.

The data samples used in this analysis were collected between June 2006 and June 2009 (Run IIb) corresponding to an integrated luminosity of 4.2 fb^{-1}. Monte Carlo (MC) simulations of various physics processes are performed using the event generators Pythia [7] or Alpgen [8] interfaced with Pythia and the detector response is simulated with Geant [9]. Candidate events must have at least one primary vertex reconstructed within 60 cm of the detector center along the beamline (z axis). The primary vertex must have at least three associated tracks, to ensure the best quality. The events are also required to have two electrons or two muons as described below. In addition, at least one jet with transverse momentum $p_{\rm T} > 20$ GeV in the detector pseudorapidity region $|\eta_{\rm det}| < 1.1$ is required in each event. A Neural Network (NN) b-tagging algorithm is used to suppress events with light-quark and gluon jets [10] and a dedicated algorithm based on a Jet Lifetime Probability (JLIP) [10] is subsequently applied to separate b quark jets from c and light-quark jets.

A. Dielectrons

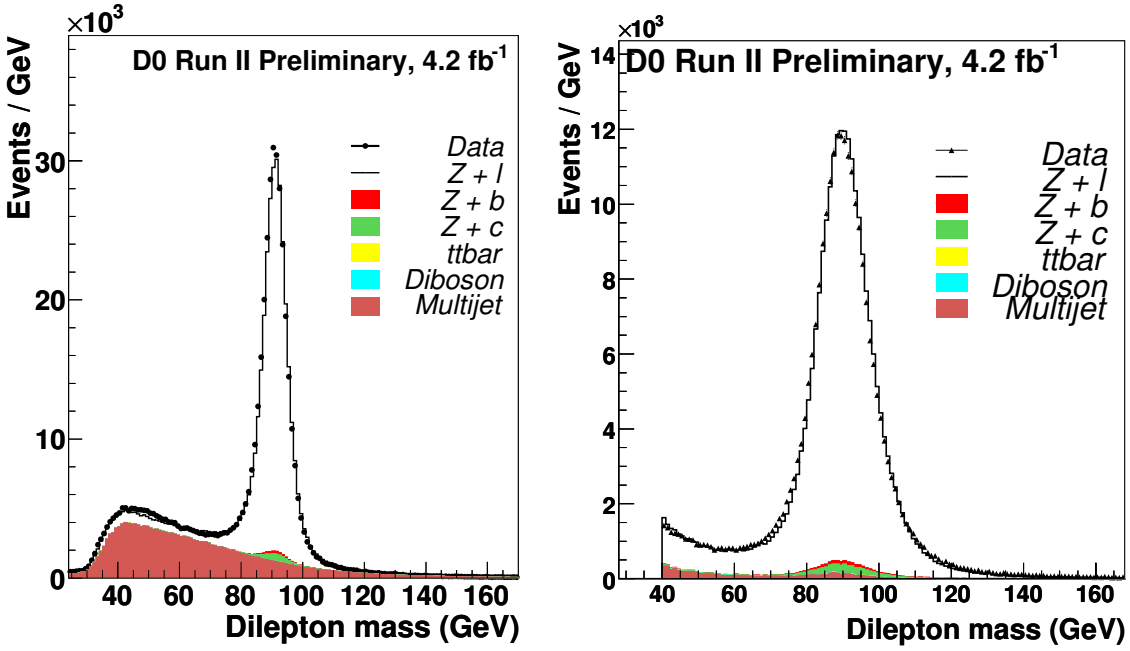
Dielectron events are required to have two electron candidates with $p_T > 15$ GeV identified in central and end calorimeters. Electrons are required to be in the $|\eta_{\rm det}| < 2.5$ and must pass selection criteria based on the calorimeter shower shape, fraction of electromagnetic energy over hadronic energy, and be isolated. In addition, any electron in the CC must match to a central track or produce an electron-like hit pattern in the tracker. The identification efficiencies for the electrons in simulations are adjusted to match those in data; corrections depend on the detector pseudorapidity $\eta_{\rm det}$ and azimuthal angle ϕ of the electron. Figure 3 (a) shows the invariant mass distribution of a pair of electron candidates in data along with various components of the background; details on the background evaluation are discussed in Section III. In the following, a Z candidate is required to have an invariant mass, M_{ee}, between 70 and 110 GeV.

B. Dimuons

Dimuon events are required to have two opposite-charge muons with $p_{\rm T} > 10$ GeV in $|\eta_{\rm det}| < 2.0$ identified with the muon and tracking detectors. To improve the momentum resolution of a muon candidate without SMT hits associated with it, the transverse momentum is corrected using the primary vertex position measured in the event. In simulations, the muon identification efficiencies are adjusted to match those in data; corrections depend on the detector pseudorapidity $\eta_{\rm det}$ and azimuthal angle ϕ of the muon. Dimuons have to pass an isolation cut to increase the likelihood that they originate from a Z boson decay. Like in the dielectron case, a Z candidate is reconstructed from two muons with the invariant mass, $M_{\mu\mu}$, between 70 and 110 GeV. Figure 3 (b) shows the invariant mass distribution of dimuon candidates in data along with various components of the background.

C. Jet Identification and Tagging

Events with a Z candidate are required to contain at least one hadronic jet with $p_T > 20$ GeV in $|\eta_{\rm det}| < 1.1$; this selection is made since the algorithm to separate b, c and light jets is currently optimized for central jets. Jets are reconstructed using the Run II jet finding algorithm [11] with a cone of radius $\Delta R = 0.5$. The energy of jets is corrected for detector response, noise, the presence of additional interactions, and energy deposited outside the reconstructed jet cone. In simulations, the jet energy is smeared to reproduce the resolution observed in data. Energy scale corrections to the jets and leptons are propagated in the computation of the missing transverse energy, $\not\!\!E_{\rm T}$. For dielectron events, all jets are required to be isolated from both electrons by $\Delta R > 0.5$. To reduce the impact from multiple interactions at high instantaneous luminosity, jets are also required to contain at least two reconstructed



(a) Dielectron Mass (b) Dimuon Mass

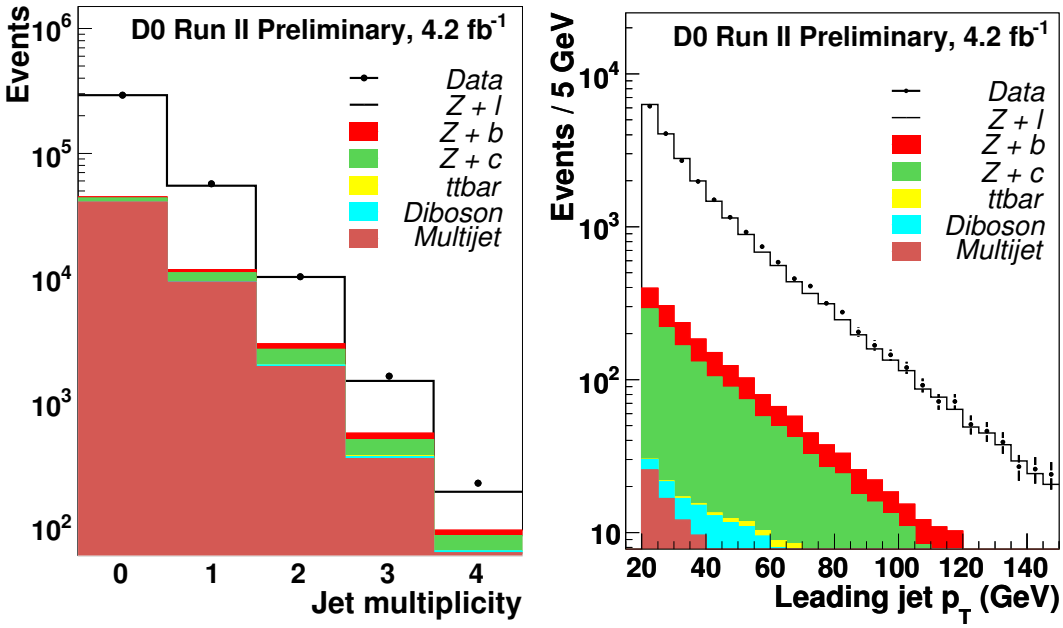
FIG. 3: The invariant mass distribution of dielectron (a) and dimuon (b) events. The data (points) are compared to the simulation (stacked histogram).

tracks that have been matched to the event primary vertex, referred to as vertex confirmation. Taggability criteria are also applied to ensure that the jet will contain enough information to be classified by the b jet tagging algorithm. Figure 4 shows the jet multiplicity and the leading jet p_T distributions in the selected $Z \to e^+e^-$ and $Z \to \mu + \mu^-$ samples, respectively, demonstrating good agreement between data and simulations.

D. Heavy Flavor Jet Candidate Selection

Before heavy flavor jet tagging, the selected Z+jets sample is dominated by light jets (light quarks or gluon jets), referred in the following as Z+ light or light jet sample. In order to suppress events with light and c quark jets, a Neural Network b-tagging algorithm is applied that exploits the longer lifetimes of B hadrons in comparison to their lighter counterparts [10]. The inputs to the NN combine several characteristic quantities of the jet and associated tracks to provide a continuous output value between zero and one. The input variables are the number of reconstructed secondary vertices in the jet, the mass of the secondary vertex, the number of tracks used to reconstruct the secondary vertex, the two dimensional decay length significance of the secondary vertex in the plane transverse to the beam, a weighted combination of the tracks' transverse impact parameter significances, and the probability that a jet originates from the primary vertex, which is referred to as the JLIP probability. The NN output value tends toward one and zero for b jets and non-b jets, respectively. In this analysis, we require at least one of the jets to have an NN value greater than 0.5. To maximize our statistics, we apply the NN selection to other jets that have $p_T > 15$ GeV if the leading jet did not pass the NN selection. After this tagging process, we use the JLIP tagger to separate the three samples.

The JLIP tagger calculates the probability that each track within a jet originates from the primary vertex based on its impact parameter information. This probability is multiplied track by track to determine an overall probability. To calculate the JLIP reduced probability, $Prob^{JLIP}_{reduced}$, a track is removed from the list of candidates that is least probable to belong to the primary vertex. This reduces the contribution of mis-measured or fake tracks on the b-tagging probability. For some cases, removing a track from a jet prohibits the JLIP algorithm from producing a sensible probability. We drop these jets and apply an efficiency to account for this loss. We define the variable rJLIP $= -\ln(Prob^{JLIP}_{reduced})$. Figure 5 shows the rJLIP distributions, normalized to unity, for different jet flavors obtained from Z+light, $Z + c$, $Z + b$ Monte Carlo samples and negatively tagged (NT) data (discussed in section IV.A). The rJLIP variable has good discriminating power between b, c and light jets, and has been previously used



(a) Number of jets

(b) Leading jet p_T

FIG. 4: (a) Jet multiplicity distribution in the dielectron channel. (b) Leading jet p_T distribution in the dimuon channel.

by DØ in the photon plus heavy flavor jets analysis [12].

III. BACKGROUND ESTIMATION

The physics backgrounds from diboson (WW, WZ, ZZ) and $t\bar{t}$ production are estimated using MC simulations with the cross sections taken from NLO calculations. The multijet background is evaluated from data. The shape of the multijet background is obtained by reversing the calorimeter shower-shape requirements in the dielectron channel and inverting the isolation requirement in the dimuon channel.

IV. ANALYSIS

Identification of heavy flavor jets and separation of light, c and b jets is carried out in two steps. First, we enrich the selected Z+jets sample with heavy flavor jets. This is done by means of the NN tagger. In the second step, we employ the rJLIP tagger to facilitate the separation between jets of different flavors and determine their fractions.

The analysis selects 28169 and 22493 candidate events containing a Z boson and ≥ 1 jet in the dielectron and dimuon channels, respectively. The background fraction in the dielectron channel is about 16%, and is dominated by the multijet production when two jets mimic isolated electrons. The dimuon channel has higher purity, with a background contribution of 0.6%. A total of 1274 (1042) events remain in the dielectron (dimuon) sample after the requirement that there be at least one b-tagged jet passing the NN output cut of 0.5. Events with large missing transverse energy (> 60 GeV) are rejected to suppress the $t\bar{t}$ background, reducing it by $\sim$50%. This requirement has minimal effect on the $Z + b/c$/jets, diboson and multijet events in the selected sample.

A. Estimation of Heavy Flavor Fractions

In order to measure the fractional contributions of different jet flavors to the final selected sample, we use a maximum likelihood fit of the rJLIP distribution in data with a combination of the light, c and b flavor jet templates. Before the fit, we subtract the non-Z+jets background contributions. A total of 922 (871) events remain in the dielectron

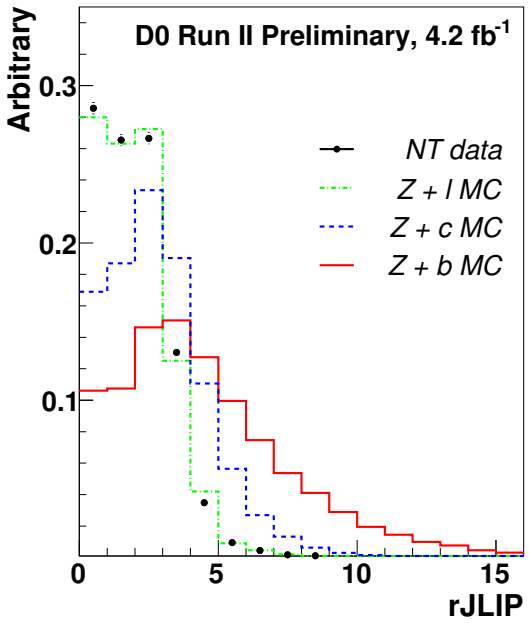


FIG. 5: Distributions of rJLIP templates for the b, c and light (l) jets.

(dimuon) channel passing all the selection requirements and after the background subtraction. The b and c jet rJLIP templates are taken from Alpgen+Pythia MC simulations with correction factors applied to account for the differences in data and MC efficiencies. The light jet template is derived from a light jet enriched data sample, referred to as negatively tagged data (NT data). NT data comprises the jets formed from tracks that have negative values for some of the inputs for the NN algorithm. We have verified that the template shapes in NT data and MC light jets agree very well, as can be seen in Fig. 5.

After generation of templates for the components of the jet sample from simulated b and c jets and the data light jet distributions, we fit the rJLIP distribution to extract the relative fractions of b, c and light jets in the tagged NN sample. The result of the light, c and b jet template fit to data is shown in Fig. 6 (a). Figure 6 (b) shows the p_T distribution of the tagged jets where the MC expectations for various jet flavors have been weighted by the fractions obtained from the rJLIP fit. The jet flavor fractions obtained from the combined dielectron and dimuon sample of 1793 events is: $Z + b = 0.191 \pm 0.030$, $Z + c = 0.384 \pm 0.072$, $Z + \text{light} = 0.424 \pm 0.054$ where the uncertainties are those from the fit due to the data statistics and finite template statistics.

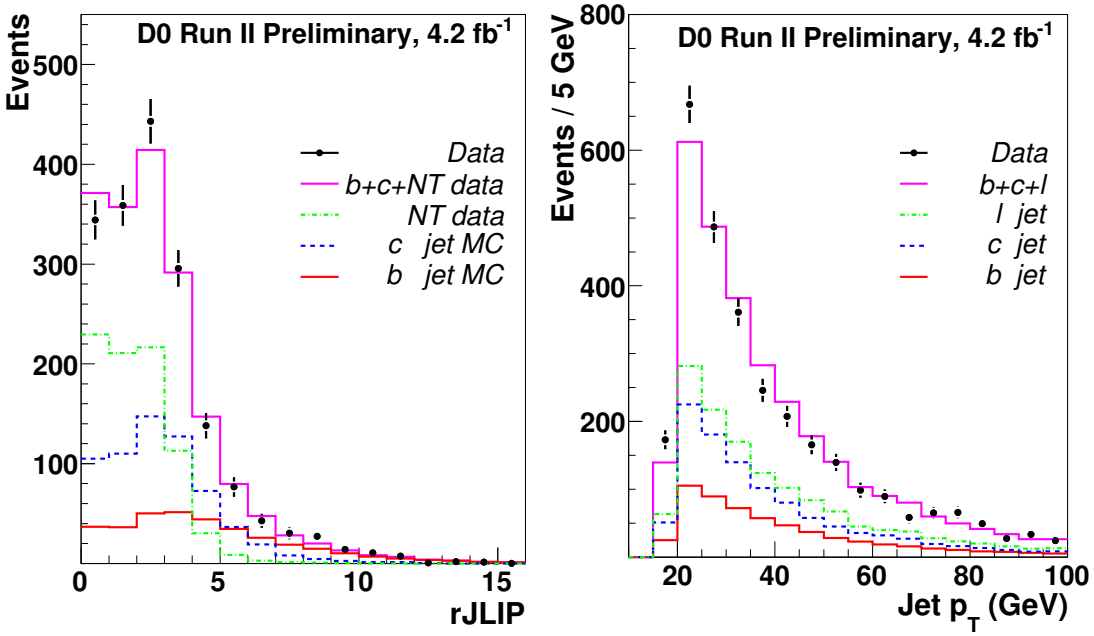
B. Determination of $\sigma(Z + b\ jet)/\sigma(Z + jet)$ ratio

The jet flavor fractions obtained from the template fitting are used to determine the $\sigma(Z + b\ jet)/\sigma(Z + jet)$ ratio as follows:

$$\frac{\sigma(Z + b\ jet)}{\sigma(Z + jet)} = \frac{P_b \epsilon_{light}^{reco}}{N \epsilon_b^{NN} \epsilon_b^{rJLIP} \epsilon_b^{tagg} \epsilon_b^{reco}} \tag{1}$$

where P_b is the number of $Z + b$ jet events obtained from the fit, $\epsilon_{b(light)}^{reco}$ is the reconstruction efficiency of b (light) jets, ϵ_b^{NN} is the b tagging efficiency of the NN tagger, ϵ_b^{rJLIP} is the rJLIP efficiency of b jets, ϵ_b^{tagg} is the taggability efficiency of b jets and N is the total number of Z+jet events before any tagging requirement.

A number of experimental uncertainties cancel out in the $\sigma(Z + b$ jet$)/\sigma(Z +$ jet) ratio measurement, such as uncertainties on luminosity, trigger, lepton identification efficiencies, etc. Several sources of systematic uncertainties have been considered in the analysis. The largest contribution, about 9.5%, comes from the uncertainty on the b jet fraction obtained from template fitting to data due to the statistics of the templates. An additional contribution due to the uncertainty in the shape of the rJLIP templates for light jets is evaluated by performing two independent



(a) rJLIP fit

(b) Leading jet p_T fit

FIG. 6: a) rJLIP discriminant distribution of observed events in the combined sample with NN > 0.5. The distributions of the b, c and light jets are weighted by their fractions found from the fit. b) The p_T distribution is shown for the same fractions.

measurements of the cross section ratio: one with the light jet template taken from the NT data and as a cross-check from the Z+light jet MC sample. The difference in the results amounts to 7.5% and is taken as the systematic uncertainty. The next largest contribution (of order 2.4%) comes from the uncertainty in the b-tagging efficiency which is fluctuated by ± 1 standard deviation around its central value and the result is propagated into the ratio measurement. Similarly, uncertainties on the jet energy scale and energy resolution are independently fluctuated by ± 1 standard deviation around their central values and the analysis is repeated to obtain a new value for the ratio. These contribute about 4%. There is an additional uncertainty in the b jet energy scale and reconstruction efficiency since the nominal jet energy scale and reconstruction efficiencies are determined with a generic data sample that mostly consists of light jets. The corresponding differences between b and light jets are evaluated with MC and propagated into the ratio measurement. The change in the ratio of 1.5% is assigned as the systematic uncertainty. Finally, uncertainties due to the background contributions are evaluated: the diboson and $t\bar{t}$ rates are fluctuated within the uncertainties on their cross sections and the multijet background is varied with the uncertainty on the normalization scale determined with data. This contribution is about 1%. Excluding the uncertainty due to template statistics, the total systematic uncertainty from above mentioned sources is 9.1%. The result for the $\sigma(Z+b\,\text{jet})/\sigma(\text{Z}+\text{jet})$ ratio in combined $\mu\mu$ and ee channels is 0.0176 ± 0.0024 (stat.) $\pm\, 0.0017$ (template stat.) $\pm\, 0.0015$ (syst.). The measurement is consistent with the theoretical prediction of 0.0184 ± 0.0022 obtained for the kinematic region considered in the analysis. The total uncertainty on the prediction includes those arising from renormalization scale, factorization scale and parton distribution functions and are 0.0015, 0.0011, 0.0011 respectively.

V. CONCLUSIONS

The ratio of inclusive cross sections for $p\bar{p} \to Z+b$ jet to $p\bar{p} \to Z$+jet production has been measured using 4.2 fb^{-1} of data collected by the DØ detector. Final states with $Z \to e^+e^-$ and $Z \to \mu^+\mu^-$ have been considered. For jets with transverse momenta $p_{\text{T}} > 20$ GeV in the pseudorapidity region of $|\eta| \leq 1$, the combined measurement of the ratio yields $0.0176 \pm 0.0024\,(\text{stat.}) \pm 0.0023\,(\text{syst.})$, which is compatible with NLO QCD predictions.

Acknowledgments

We thank the staffs at Fermilab and collaborating institutions, and acknowledge support from the DOE and NSF (USA); CEA and CNRS/IN2P3 (France); FASI, Rosatom and RFBR (Russia); CAPES, CNPq, FAPERJ, FAPESP and FUNDUNESP (Brazil); DAE and DST (India); Colciencias (Colombia); CONACyT (Mexico); KRF and KOSEF (Korea); CONICET and UBACyT (Argentina); FOM (The Netherlands); STFC (United Kingdom); MSMT and GACR (Czech Republic); CRC Program, CFI, NSERC and WestGrid Project (Canada); BMBF and DFG (Germany); SFI (Ireland); The Swedish Research Council (Sweden); CAS and CNSF (China); and the Alexander von Humboldt Foundation. We would also like to thank John Campbell for the useful suggestions on the evaluation of theoretical prediction using MCFM.

[1] J. M. Campbell, R. K. Ellis, F. Maltoni, and S. Willenbrock, Phys. Rev. D **69**, 074021 (2004).
[2] F. F. Cordero, L. Reina, and D. Wackeroth, Phys. Rev. D **78**, 074014 (2008).
[3] V. Abazov *et al.* (D0 Collaboration), Phys. Rev. Lett. **94**, 161801 (2005).
[4] T. Aaltonen *et al.* (CDF Collaboration), Phys. Rev. D **79**, 052008 (2009).
[5] V. Abazov *et al.* (D0 Collaboration), NIM A **565**, 463 (2006).
[6] R. Angstadt *et. al.*, "The Layer 0 Inner Silicon Detector of the D0 Experiment", To be published in NIM A, arXiv:0911.2522.
[7] T. Sjöstrand *et al.*, Comput. Phys. Commun. **135**, 238 (2001) versions 6.319 and 6.413.
[8] M. Mangano *et al.*, JHEP **07**, 001 (2003), version 2.11, `http://mlm.web.cern.ch/mlm/alpgen`
[9] R. Brun *et al.*, GEANT Detector Description and Simulation Tool, *CERN Program Library* **W5013** (1994).
[10] V. Abazov *et al.* (D0 Collaboration), To be published in NIM A, arxiv:hep-ex/10024224.
[11] G. C. Blazey *et al.*, in *Proceedings of the Workshop: "QCD and Weak Boson Physics in Run II,"* edited by U. Baur, R. K. Ellis, and D. Zeppenfeld, (Fermilab, Batavia, IL, 2000) p. 47; see Sec. 3.5 for details.
[12] V. M. Abazov *et. al.* (D0 Collaboration), Phys. Rev. Lett. **102**, 192002 (2009).
[13] J. Campbell, R. K. Ellis, Phys. Rev. D **60**, 113006 (1999). http://mcfm.fnal.gov/.